

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
8 February 2001 (08.02.2001)

PCT

(10) International Publication Number
WO 01/10125 A1

(51) International Patent Classification⁷: H04N 7/173

(21) International Application Number: PCT/US00/21127

(22) International Filing Date: 2 August 2000 (02.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/146,893 2 August 1999 (02.08.1999) US
09/475,807 30 December 1999 (30.12.1999) US

(71) Applicant and

(72) Inventor: CHUNG, Randall, M. [US/US]; EdgeStream Inc., 22921 Triton Way, Suite 231, Laguna Hills, CA 92653 (US).

(74) Agent: YU, Philip, K.; Registered Patent Attorney, 100 E. Huntington Drive, Suite 100, Alhambra, CA 91801-1022 (US).

(81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

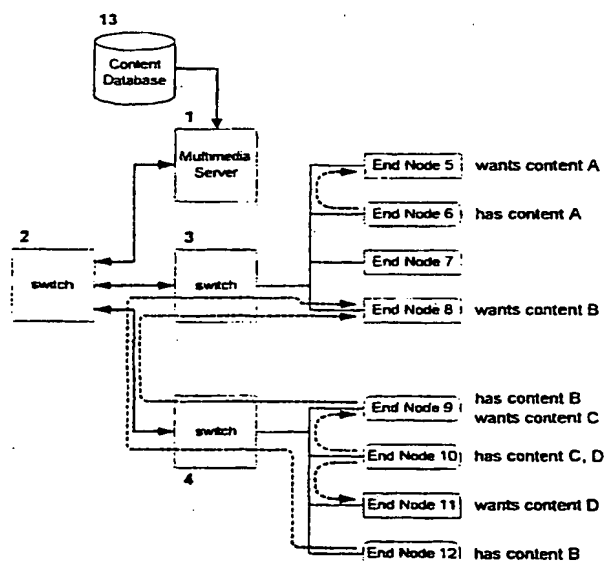
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

- With international search report.
- Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: VOD FROM A SERVER OR A USER TO ANOTHER USER



(57) Abstract: A system and method for delivering data to multiple end users (5-12) using distributed storage system connected through a data communications network. The end user equipment for audio or video playback or recording contains a storage unit, such as a magnetic disk. Each storage unit can store data, such as digital video content, i.e. movies, or digital audio content, i.e. songs. The storage unit can also be configured to store portions, or fragments, of digital video or audio content, where the storage location, playback, and recording of digital video or audio content of each storage unit is administered, managed and controlled by a central facility (1). The content for playback of a video or audio content, or portions thereof, may reside on one or more physically separate storage unit, or may also reside on a central storage unit (13).

WO 01/10125 A1

VOD FROM A SERVER OR A USER TO ANOTHER USER

5

Cross Reference to Related Application

This application claims priority from provisional application Serial No. 60/146,893, "Method and Apparatus for Data Delivery Using Distributed Storage System," filed August 2, 1999.

10

Field of the Invention

The present invention relates to data communications networks and more specifically to a method and apparatus for delivering data using distributed storage units through a network, such as the Internet.

15

Art Background

Ever since the advent of VCR's, there has been a public desire to have a movie-on-demand system to alleviate the need for the viewers to go to the video store. A "video on demand" ("VOD") system provides viewers the ability to select an arbitrary movie or video program and then view the video content at a requested time, whether immediately, a few hours later or even a few days later. The delivery of the video is purely electronic, so that physical media like video tapes or video discs need not be purchased, nor rented at a video store and then returned after viewing.

20

Cable companies have implemented some form of video delivery system for their cable customers. However, due to limitations in the storage space and bandwidth, the selection of movies is quite limited and the viewers can only watch the movie at the time designated by the cable company. For those who feel like watching an old *Godfather* movie on Wednesday afternoon, they are out of luck. The system offered by the cable company is therefore not a true VOD system, much less a system with any degree of flexibility suitable for a broad range of viewers.

30

Implementing a practical video on-demand system for a large number of viewers is difficult and expensive. It typically requires a multimedia server which can deliver multiple simultaneous continuous streams of audio or video content to hundreds or thousands of different end customers. Ideally, the customers could choose their content from a library
5 of hundreds of movies or songs. However, a typical hard disk drive may only be able to hold a few full length movies, which means that a multimedia server would have to contain 50 or more disk drives. Figure 1A illustrates such a conventional cable broadcasting system for "video-on-demand" provisioning.

Another difficulty in implementing a large shared multimedia server is the need to
10 handle and transmit multiple streams of digital multimedia content. A standard definition compressed television signal would be transmitted at an average data rate of about 4 megabits per second. If one multimedia server had to serve one hundred different programs simultaneously, the average total bandwidth required would be 400 megabits per second. This large amount of data would also have to be transmitted to the one hundred
15 viewers without appreciable discontinuity, or the viewing experience would be ruined. A network to support such a high data rate with a guaranteed Quality of Service (QoS) to many users from a central server is thus very expensive, and has been impractical for anyone to build.

With the advent of communication through the Internet using cable modems, it has
20 been suggested that cable modems may enable the provisioning of video on-demand by cable companies. It has been hoped that cable modems coupled with the Internet can achieve what cable companies have failed, that is, a video on-demand system using the Internet.

Figure 1B illustrates how video on-demand might be provisioned over a cable TV
25 broadcast system with cable modems for a metropolitan area. In this system, analog TV signals are received by a satellite dish 1. The analog TV signals are combined with video on-demand signals and digital communication signals from the Internet through a signal combiner 10.

In Figure 1B, the video-on-demand signals are digital, although they can also be
30 analog signals. Digital content, such as World Wide Web pages, on the Internet public network 2, are connected to the system through a router 3. The router 3 is used to

translate packet data from one network to another, such as between the Internet network and the example Ethernet network switch 4. The Ethernet network switch 4 directs data to and from the Internet router 3, the fiber return channel demodulator 5, a file server 6, the multimedia server 8, and the modulator 9. The fiber return channel demodulator 5 translates signals returning from the customer premises equipment ("CPE") through the optical fiber 11, back into digital format, where it might then be sent to the Internet. A file server can be used to hold World Wide Web pages or for caching Internet information. The multimedia server controls the content database 7, and can read out multiple movies or songs simultaneously. The content database holds all of the digitized movies or songs. Digital data from the multimedia server or the Ethernet switch going to the end users can go through a digital to analog modulator 9, before being combined with the analog TV signals from the satellite dish 1.

Referring still to Figure 1B, signals going between the optical fiber 11 and the coax cables 13 can be translated by the coax-to-fiber converter 12. The analog TV signals would be frequency converted through the set top box 16, for transmission to the TV set 17, while the digital signals would be converted to digital by the cable modem 14 for use by the PC 15. The customer premises equipment 18 (CPE) at a single residence may contain one set of equipment, while another residence may contain another set of CPE 19, with each customer viewing different video, audio, or data content.

It is not unusual to have hundreds or thousands of customers served by one set of provider equipment. That means a multimedia server would have to potentially serve different movie or music data streams to each customer simultaneously. This is practically impossible if there are thousands of customers and only one multimedia server. One solution may be to have multiple multimedia servers, with each one serving a subset of the customers; however, this also multiplies the cost of servicing all of the customers. A remaining problem is that a very large amount of multimedia data must travel from a centralized multimedia server, or group of servers. Because each customer may be viewing different videos at different points in time (for instance, at the beginning, middle, or end of the video) from the multimedia server, each video is treated as a separate channel. The bandwidth of hundreds or thousands of separate channels, in addition to the normal broadcast channels is higher than what current transmission pathways can handle today.

Therefore, it is desirable to have an on-demand data delivery system where the multiple recipients can receive the same or different content at any time they choose.

It is also desirable to have such an on-demand data delivery system without exhausting the bandwidth of the server or the transmission infrastructure.

Summary of the Invention

A system and method for delivering data to multiple end users using distributed storage system connected through a data communications network is disclosed, where the end user equipment for audio/video playback or recording contains a storage unit, such as a magnetic disk. Each storage unit can store data, such as digital video content, i.e. movies, or digital audio content, i.e. songs. The storage unit can also be configured to store portions, or fragments, of digital video or audio content, where the storage location, playback, and recording of digital video or audio content of each storage unit is administered, managed and controlled by a central facility. The content for playback of a video or audio content, or portions thereof, may reside on one or more physically separate storage unit, or may also reside on a central storage unit.

The content may also be read out and transmitted concurrently from one or more physically distributed storage units to one or more other physically separate sites for playback, where an entire piece of content, such as a movie, may be separated into equal or unequal sized pieces, and one or more pieces may be stored on multiple physically separate storage units. The assignment of the storage locations of the pieces, the order of transmission, priority of transmission, and destination of the pieces is centrally administered and controlled. The content can also be transmitted simultaneously to storage units other than the one needing it for display, to anticipate the need to deliver the same content to other customers in the near future. This allows for dynamic allocation of resources at any time, peak or off-peak, in anticipation of heavy downloading of certain programs, such as new releases.

Further, the data communications network connecting between the storage units can be either wired or wireless. While the Internet is used with the cable modems, other forms of networking connections can work just as well.

As will be appreciated by those skilled in the art, what is disclosed in the following description is a "set-top box" which is coupled to a disk drive, or a file server, either

locally or remotely, where the "set-top box" acts as a file server for other "set-top boxes." As such, the set-top boxes in the aggregate can serve as a massive and yet distributed storage facility for each of the member set-top boxes. The stored programs or files may also be fragmented into pieces of arbitrary sizes for distributed storage or transmission over multiple set-top boxes. Also, what is disclosed is a personal computer which contains a disk drive which can act as a file server for other computers or set-top boxes, where the files may be fragmented into pieces of any arbitrary sizes for distributed storage or transmission over multiple computers or set-top boxes.

Brief Description of the Drawings

Figure 1A shows a conventional video-on-demand system over a cable broadcast system.

Figure 1B shows an exemplary video-on-demand system over a cable TV broadcast system, where users end-users are equipped with cable modems.

Figure 2A illustrates an example of how a single multimedia server might serve AV content to a hundred users simultaneously.

Figure 2B illustrates an example of how three cache memories can be inserted in between the multimedia server and the customers.

Figure 3A is a simplified block diagram of an End Node Equipment at a user's site.

Figure 3B depicts a simplified block diagram of a conventional VOD system with a centralized server and end node equipment at user's site.

Figure 4A is a simplified block diagram of an End Node Equipment with cable modem for use with a TV in accordance with the present invention.

Figure 4B is a simplified block diagram of an End Node Equipment with cable or DSL modem for use with a PC in accordance with the present invention.

Figure 4C depicts a simplified block diagram of a distributed server system of the present invention.

Figures 5A, 5B illustrate simultaneous transfer of content from multiple end nodes to reduce playback starting time.

Figure 6 depicts a simplified block diagram of a VOD system based on DSL network.

Figure 7 is a flow diagram depicting the process of distributing content to end node units.

5 Figures 8A and 8B illustrate another embodiment of the VOD system with the content being delivered from a combination of end nodes and content servers.

Detailed Description of the Preferred Embodiment

10 A method and apparatus for data delivery using distributed storage system is disclosed. In the description that follows, the present invention is described in terms of servers, switches, routers, databases, end nodes, networks and the Internet, which are the terms typically used by those skilled in the art in communicating among themselves. Also, it should be understood by those skilled in the art that the term "data" can represent audio or video content, multimedia content, or simply binary bits transmitted across a communications network.

15 As previously mentioned, one possible way of reducing the multimedia server's bandwidth requirements is to have storage units along the main network transmission paths to cache the first transmission of the content. Future requests for the same content from other end users can be serviced from the caching storage unit.

20 Figure 2A shows an example of how a single multimedia server might be requested to serve AV content to a hundred users simultaneously. Many customers may be viewing the same AV content, because some content is more popular than others. For instance, a newly released movie may be much more popular than one several years old. However, it is desirable to allow each customer to start viewing the movie at a time of their own choosing, rather than on a fixed schedule, which happens to be the way cable companies
25 offer video programs. This means that although the same movie is being viewed by more than one customer, the data must be sent in duplicate, because the movies may be viewed at different points in time.

30 Figure 2B shows how cache memories may be inserted in between the multimedia server and the customers to ease the delivery demand. Each cache memory serves a smaller group of the full set of customers. If AV content has not been previously stored in

a group's cache memory, it will be provided by the central multimedia server. The AV content will be transferred to the customer, and also temporarily stored in the group cache memory. If another customer in the group requests the same AV content, and the AV content still resides in the cache memory, then the cache memory can readily provide the content. This reduces the load on the central multimedia server and reduces the bandwidth used by the server. However, there is an obvious problem to this approach. If the cache memory is too small, then only a small portion of the AV content can be stored, causing many requests for content to bypass the cache. This reduces the effectiveness in reducing the bandwidth and load on the central multimedia server. However, making the cache memory large will increase the cost.

As will be disclosed, the present invention is directed to providing a solution to the problems with the conventional ways of distributing data such as digital multimedia by having a system of storage units at the end nodes, acting as the primary storage of the multimedia content. A centralized multimedia server could still be used to augment the delivery of multimedia content, but the bulk of the requests for multimedia data from the end nodes will be serviced from the individual storage units at the other end nodes.

Such a distributed server system can address the problems inherent in the current centralized server system when used for distributing time-critical multimedia content. As will be described below, Figures 3A and 3B illustrate the conventional system utilizing a centralized server, while Figures 4A, 4B and 4C illustrate a distributed server system in accordance with the present invention.

Figure 3A is a simplified block diagram of a conventional End Node Equipment which can be installed at the user's site. It is generally a set-top box ("STB") coupled to the TV. The STB generally has a cable modem and a digital video and audio decompression circuitry. As shown in Figure 3B, the multimedia server 1 controls the content database 13, which stores all of the multimedia content. The network switch 2 controls the data transfers among the server, network switch 3, and network switch 4. Network switch 3 switches the data from network switch 2 to the user end nodes 5 through 8, while network switch 4 switches the data from network switch 2 to the user end nodes 9 through 12. In this example, end nodes 5, 8, 9, and 11 are receiving multimedia data from the centralized server 1. The server 1 and switch 2 may be separated

by a long distance, possibly requiring expensive high bandwidth connections. The content data streams A and B for end nodes 5 and 8 are routed from switch 2 to switch 3, while the content data streams C and D for end nodes 9 and 11 are routed from switch 2 to switch 4. Switch 3 then routes the data it receives to end nodes 5 and 8, and switch unit 4 routes the data it receives to end nodes 9 and 11. In this system, all four of the data streams A through D must come from content database 13, through server 1, through switch 2, and then through switches 3 and 4. Because of the time critical nature of the multimedia data, all of the components and network interconnections between the content database 13 and the end users must be very high speed, with low latency. The system as in Figures 3A and 3B clearly will have its limitations, particularly when there are a large number of end nodes requesting data simultaneously.

Figure 4A shows a consumer-oriented End Node equipment for viewing digital video on a TV. Its STB contains the cable modem, a storage unit such as disk storage and digital video and audio decompression circuitry. Of course, there is the TV for displaying the program from the cable. Figure 4B shows PC-oriented equipment for viewing digital video, using either a cable modem or DSL modem for getting the data to the PC with a storage unit, e.g. disk storage.

For a PC-based end node equipment as shown in Figure 4B, its operation can be controlled by software written in Sun's Java Media Framework (JMF), which can support several Internet protocols called Real Time Protocol (RTP) and Real Time Streaming Protocol (RTSP). These software programs are generally considered "off the shelf" by those skilled in the art and their capabilities are also well understood by those skilled in the art. Java programs can also be written to use the JMF library to facilitate the transfer of data, which can be programmed to support its transmitting end node and another program to support its receiving end node. Those skilled in the art will also appreciate that software may be written for a PC to control the coordinating tasks.

Currently, for a demonstration and prototype unit, the person running the receiving PC installs the Sun Java Development Kit (JDK) software, available for download at <http://www.javasoft.com>, and the person can then obtain the receiver software, e.g. via electronic mail from the service vendor.

In a production version of the software for operating the end node equipment, an end user can visit the vendor's web site, to install the receive/transmit software on-line. An auto-start software routine can also be installed in the PC's startup directory so that the receive/transmit software can start up when the PC is turned on or rebooted. The installation program can also store personalization information in a special file, to identify the PC and the current PC user, since several people may share a single PC (this special file is commonly known as a "cookie." file).

The following illustrates how an end node equipment may be initialized at start-up. It should be noted that those skilled in the art can readily devise their own initialization routine and protocol while developing their distribution systems in accordance with the present invention. The following discussion is for illustration purposes only. At start-up, the receive/transmit software resident at the end node equipment, e.g. PC, can send a message to the control software on the service provider's web site, acknowledging that it is ready to receive or transmit if requested, as well as sending the identification information. Upon receipt of the acknowledgement, the service provider's control software knows that that particular PC is available.

Once the service provider's server knows that an end unit is available, it can send commands and data to:

- add new content with the end unit receiving command and data;
- delete old content with the end unit receiving a command;
- transmit existing content with the end unit receiving a command;
- check existing content for tampering or perform other maintenance tasks.

In one embodiment, the control software may need to know when a particular PC is no longer available. This scenario may happen when the user shuts the PC down, when the PC crashes, when the user's power fails, or when the cable network simply stops working. If the user shuts the PC down smoothly, i.e. not abruptly, the control software at the PC can be notified of the shutdown, and can send a message back to the server that it will no longer be available.

If the PC crashes or its availability is removed without notification to the provider's server, the distribution system should preferably recover from that without ruining the viewing user's experience. If a series of transfers is being setup, say because a viewer at

another end node unit wants to watch a movie, then the server can immediately check, e.g. poll, with all of the end units that it thinks are available. If any end node unit fails to respond, then the server could mark that end node unit as unavailable, and substitute another end node unit. The initial transfer list would then always start out with end node
5 units that have been confirmed to be available.

Once the movie starts, the receiving PC can be configured to expect data from each transmitting PC by a certain time. If one PC doesn't deliver data by the scheduled time, due to a possible crash, then the receiving PC can notify the provider's server. The server can then immediately substitute another end unit, by giving the substituted end unit
10 the remaining part of the transfer list, and by notifying the receiving unit of the substituted unit's internet address. To ease the transition, the provider may keep a backup server which would provide the delayed content to the receiving end unit for a few seconds, until the substituted end unit is up to speed.

Reference is now to Figure 4C. As shown in Figure 4C, the content database 13 is
15 set up so that it can offload some of its duties to the whole distribution system. All of the multimedia data is instead distributed among the storage units local in each of the user end nodes (not shown). It may not be necessary for all of the user end nodes to have local storage units, but system performance is improved when more and more user end nodes have storage units and participate in the distribution. Multimedia server 1 now acts as a
20 centralized administrator, keeping track of how the multimedia content is distributed among the user end nodes. The network connection between server 1 and switch 2 now needs much less bandwidth, because only commands and status information are transferred, and not the multimedia data, the storage of which is now distributed among the end nodes. As will be appreciated by those skilled in the art, the storage unit at each
25 end node can be a hard disk drive, or any kind of storage device capable of storing, transferring and playing back its multimedia content. As the prices of the storage devices plummet due to technological maturity, this form of distributed storage becomes more attractive and cost effective.

In the system exemplified in Figure 4C, end node 5 wants multimedia content A,
30 end node 8 wants multimedia content B, end node 9 wants multimedia content C, and end node 11 wants multimedia content D, while the storage unit of end node 6 contains multimedia content A, the storage unit of end node 9 contains multimedia content B, the

storage unit of end node 10 contains multimedia content C and D, and the storage unit of end node 12 contains multimedia content B.

The distributed storage system, as illustrated in Figure 4C, can be initialized by the multimedia server 1 directing various contents, e.g. movies, to be transferred from the content database 13, or wherever contents may be available, to the end nodes. The distribution can be based on the anticipated popularity of certain contents and whether the end nodes are within the same switching network. For example, if there are 200 end nodes sharing the same immediate switching network, e.g. in the same neighborhood, a newly released movie may need to be resident at, say, 10 end nodes to provide assured delivery. The delivery can also come from end nodes at other not-so-immediate switching networks, if the delay is not significant. The other 190 end nodes can be populated with other contents, in fragments, in whole or in combination thereof. This is only the initial phase, since as the demands change, the location and distribution can certainly be changed to reflect the usage, which can be tracked using common tools for statistical analysis.

Moreover, the initialization or subsequent location management can be done during off-peak hours, since no intervention from the consumer should be needed. Therefore, at 3 A.M., while most consumers are asleep, the storage units at their end nodes may be quietly receiving instructions, preferably from the central server, transferring contents, or updating the central server. As new contents are released, the new ones will cause the not-so-new contents to be deleted or replaced, either gradually or all at once. The distributed storage system is thus dynamic and able to continue to adequately support the consumers' ever-changing needs.

Because storage of the content is distributed, end node 6 can act as a server for end node 5. If the network connection between the two nodes is a shared medium, then it may be possible for end node 6 to transmit the content A data directly to end node 5, otherwise it can transmit the content A to end node 5 by sending it to switch 3, which will then send it to node 5, using conventional Internet, or any data communications networking, addressing techniques. The multimedia server 1 can preferably act as a "traffic controller" to receive requests from end node 5, check availability of the requested content, determine the most expedient way of transfer, e.g. from end node 6, and command end node 6 to transfer the content A to end node 5.

Still referring to Figure 4C, it may happen that when end node 6 is responding to a request for content A from end node 5, the viewer at end node 6 also requests such content A for playback. In such situation, data can be read out of the storage unit, e.g. hard drive, into the processor's memory, whether it is the PC or the STB. Then the data is transferred out of the processor's memory and out to the cable or DSL network. If the same data is needed at very close to the same time for local viewing, that data is likely still cached in the memory and can be sent to the local display. If it isn't still cached in the memory, then the data can be re-read from the storage unit, e.g. hard drive. As can be appreciated by those skilled in the art, current PC systems with their high speed processors and communications functionality, as well as the STBs, can readily handle two sets of read requests simultaneously.

An alternative to the multimedia server acting as a centralized traffic controller is to let each end node broadcast its request to other end nodes. The end nodes with the requested content will monitor the transmission status to determine if it is the best source for the requested content. If so, it will reply and transmit the requested content to the requester. When there are multiple replies to its request, the requesting end node will determine the priority for receiving and notify the source end node(s) of the priority for transmission. If transmission is bogged down, the requesting node can go to the next source end node according to the priority. The second source end node can transmit to pick up where the first source end node leaves off.

Similarly, end node 9 can act as a server for end node 8, providing it with content B, and at the same time, receives content C from end node 10. End node 10 stores content C and D, and can act as a server by sending content C to end node 9 and simultaneously sending content D to end node 11.

The end nodes do not have to be servers by themselves only. They can work, sequentially or concurrently, with other end nodes to complete a delivery. End node 8 can receive content B from both end node 9 and end node 12. This may be desirable to reduce the transmission time of content B, particularly if the upload (direction from the end node to the switch) speed is lower than the download (switch to end node) speed. This is often the case for cable modems and ADSL (Asymmetric Digital Subscriber Line) network technology. It also can be useful in the situation where the playback of the content starts before the entire content has been downloaded to the end node. For instance, the playback

of content B may have started because it was expected that the download from end node 9 only would complete before the end of the content was reached, based on the speed of the early part of the download from end node 9 to end node 8. However, once playback starts, if the download speed is reduced due to unexpected network congestion from switch 4 to switch 3 or due to unexpected reduction in upload speed from end node 9 to switch 4, then the unviewed portion of content B can also be transferred from end node 12 to end node 8.

By way of an example, suppose that end node 8 has received the first 25% of content B from end node 9. Based on the download rate of this first 25% of the content as well as the playback rate for the content, end node 9 starts the playback. In this example, the playback rate is equal to the download rate. At the 30% reception point, the download rate slows down by a factor of 2. If the download rate is not increased, the playback of content B will be interrupted at the 90% point of the movie or song. At, say the 35% point, end node 12 can be commanded to immediately start transferring the last 20% of content B, and end node 9 can be commanded to continue transferring the content from the 35% point but to stop at the 80% point. This method can be useful where the upload transfer rate or the switch to switch transfer rate is lower than the download transfer rate.

The method of simultaneous, or concurrent, transfer from multiple end nodes can also be used to reduce the playback starting time for a network where the playback data rate and the download data rate are higher than the upload data rate. For instance, suppose the playback rate is 2 megabits per second, the download rate is capable of supporting 3 megabits per second, and the upload speed of a single end node is 1 megabit per second. For the network shown in Figure 5A, end node 5 wants to play content A, which is stored as a number of different fragments in end nodes 6 through 15. The information on the location of the fragments at the different nodes can preferably be stored on the multimedia server 1. The upload and download capability of the different nodes can also be stored on the multimedia server 1, or there could be an immediate test transfer to determine the upload and download speeds.

Referring to Figure 5A, end node 5 would contact the multimedia server 1, which would coordinate the transfer of the fragments to end node 5 from the other end nodes. In the first phase, end nodes 6, 7, and 8 would transfer their stored fragments to end node 5, which would receive the three fragments concurrently at 3 megabits per second, even

though the individual fragments from each transmitting end node can be uploaded only at 1 megabit per second. Since the fragment containing 0-5% of the content from end node 6 is being transmitted at half the playback speed, end node 5 could not start playing the content until the first half of the fragment has been received. If it started playing any
5 sooner, the data would run out before the fragment could be fully received, causing a gap in the playback.

It may occasionally happen that the Internet address of the end nodes is not fixed, but can change from time to time. A program on each of the end nodes can be used to inform the controlling multimedia server 1 when its Internet address changes.

10 When the fragments from end nodes 6, 7, and 8 have been transferred to end node 5, the multimedia server 1 could arrange the transfer of the next 15-30% fragments from other end nodes containing the necessary fragments, such as end nodes 9, 10, and 11, as shown in Figure 5B. While end nodes 9, 10, and 11 are transmitting data to end node 5, end nodes 6, 7, and 8 would become available to transfer data to a different end node, if
15 needed.

It should be noted that while only one content database and one multimedia server is depicted in Figure 5B for illustration purposes, the entire distribution system in accordance with the present invention may implement multiple content databases with multiple multimedia servers, scattered throughout the county, state or country. Content
20 going to one end unit may come from one or more content servers as well as from multiple other end units. This "aggregation" effect can offload the burden on any one particular server or database, while providing an efficient scheme to achieve content delivery from databases and servers in proximity of the end units.

Referring to Figure 5B, when end node 5 has completed viewing content A, its
25 storage unit will have a complete copy of all of the fragments of content A. If content A is expected to be popular with other end nodes, e.g. new releases, the entire content could remain stored on end node 5, so it could be delivered to other end nodes, or just a few portions could be kept. What portions to be kept at this end node can be determined based on the storage, or distribution, situation of content A at other end nodes. For
30 example, the storage situation may be such that only the first 30% fragment of content A is needed, since other end nodes may contain sufficient other fragments. If content A is not

expected to be viewed in the near future, it could be deleted immediately, or designated to be the first content to be deleted when other content needs to be stored in end node 5's storage unit.

Reference is now turned to Figure 7, where a flow diagram depicting the process of distributing content to end node units as directed by the server is described. As shown in Figure 7, end-user unit (EUA1) notifies the central server it is on-line, active and ready, for example, after the EUA1 is first installed at an end-user's house. Upon such notice, the central server transmits data block N to EUA1 for storage. When EUA2 requests data block N from the central server, the central server will notify EUA2 that data block N will be transmitted from EUA1. The central server can thus command EUA1 to transmit the requested data block to EUA2. At this point, the Internet transmit address is EUA1's, while the Internet receiver address is EUA2's. Upon receipt of the requested data block, EUA2 notifies the central server that transmission is now completed.

Reference is now to Figures 8A and 8B, where another embodiment of the VOD system in accordance with the present invention is shown. It should be pointed out that in the system previously shown in Figures 5A and 5B, a number of end nodes are served by a single multimedia server 1. Figures 8A and 8B show an alternative to, or elaboration of, this system, where the multimedia server's function has been split, or essentially off-loaded, into separate parts, and the content is delivered from a combination of end nodes and multimedia content servers. The content delivery is now accomplished by distributed content servers as well as by other end nodes, or the combination of servers and end nodes.

In this system as shown in Figures 8A and 8B, the end nodes C and D may be personal computers (PC's) or other equipment which normally reside in a home. Control Server A may be an Internet server computer which performs the control and coordination role. An example of this Internet server computer is a Compaq Proliant server or its equivalent, using the Microsoft NT operating system. Web Server B may be an Internet server computer, an example of which is Sun Microsystems' Enterprise server or its equivalent, using the Sun Solaris operating system, which delivers web pages to end nodes, but which may or may not hold the multimedia content. Multimedia Content Servers E and F may be Internet server computers which hold multimedia content. Although the functions are separated for this example, it is possible for an Internet server

computer to perform and combine the control, web serving, and multimedia content serving functions in any combination, as can be appreciated by those skilled in the art.

Figures 8A and 8B show another embodiment of the VOD system, in accordance with the present invention, delivering multimedia data to an end node from a combination of multimedia content servers and end nodes. This embodiment thus makes it possible to build a system which delivers multimedia data to an end node, or multiple end nodes, only from multimedia content servers.

As previously described in the present application, the multimedia content, such as a movie file, may first be fragmented into smaller files, and the smaller files may be distributed for storage by end nodes and/or other servers. For this example, the content fragments would be distributed onto End Node C (step 3), End Node D (step 4), Multimedia Content Server E (step 3), and Multimedia Content Server F (step 3). The present invention also makes it possible to distribute copies of the file fragments to additional end nodes or servers, so that each file fragment may reside on more than one end node or server. This redundancy can allow the substitution of one end node or server for another, which may be useful if an end node or server is slow in transmitting data. The data may reside on these end nodes and servers, and may be transmitted when needed.

Also, as previously described, the viewing end node may contain some software which may re-integrate the file fragments into the original file order. Finally, the end nodes or Internet servers which hold the movie file fragments may contain some software which allows it to accept commands from a control server and to deliver the file fragments to an end node or to an Internet server.

In Figure 8A, step 1 occurs when a user at End Node D, say, wishes to view a movie. It sends a request to Web Server B for the movie data. In step 2, Web Server B passes the request to Control Server A. Control Server A then determines which end nodes or Internet servers contain the desired movie file fragments. In step 3, Control Server A sends instructions to End Node C, Multimedia Content Server E, and Multimedia Content Server F to deliver the desired movie file fragments to End Node D. In step 4, shown in Figure 8B, End Node C, Multimedia Content Server E, and Multimedia Content Server F together deliver the movie file fragments to End Node D. It should be pointed out that the re-integration, or re-assembly, of the various fragments can be done by the viewing end node, if each fragment is coded to provide unique identification with respect

to other fragments from the same title. Such re-integration or re-assembly may also let the viewing end node obtain a "script" of the fragments to be delivered from the server responding to the viewing end node's initial request. This "script" will facilitate the re-integration and re-assembly of the fragments, much like an instruction manual on how to
5 assemble a bicycle.

It should be pointed out that once a movie title, or a data file, is fragmented, each fragment should have a unique code identifying its position in the full length movie. Preferably a standardized way of position-coding the movie or data file is established, in a similar manner Internet files are packetized and delivered separately throughout the
10 Internet, only to be reassembled at the last point for the reception.

Although it is not shown, it should be apparent to those skilled in the art that it is possible for Web Server B to also act as a multimedia content server.

It should be noted that each fragment of data being stored by an EUU can be encrypted with a different encryption key for security reasons. For example, if an EUU
15 stores 5 fragments from 5 programs, i.e. 1 fragment from 1 program, each fragment can have its own encryption key. Having the fragments or programs encrypted will prevent any unauthorized access of the EUU by its user. As can be appreciated by those skilled in the art, the storage in each EUU is preferably transparent to the user, whether the EUU contains a full-length feature or a collection of fragments of various contents.

20 In accordance with the above disclosure, the present invention provides continuous playback of content, with a minimal amount of waiting, even though the upload transmission rate of content from a single end node is lower than the playback and download rates of the playing end node by fragmenting the content and distributing the content among a large number of networked storage nodes. The present invention also
25 allows multiple end nodes to view the same content, with a small delay between each viewing, where the delay is set by the time duration of a fragment. The present invention can be used with any means of two way transmission of data. For instance, the network could use DSL (digital subscriber line) technology shown in Figure 6, instead of, or in conjunction with, the cable TV technology shown in Figure 1.

30 While the present invention for data delivery has largely been described in terms of implementing a video-on-demand system with distributed storage, it should be appreciated

by those skilled in the art the present invention can also implement a "data-on-demand" system. In such a system, the storage unit at each end-user's site will store data for access by other users or by itself. The stored data can also be portions, or fragments, of distinct kinds of data, which can be assembled with other portions stored elsewhere to form a complete piece, provided there is mechanism, e.g. via a central server, to control, track and manage the storage, transmission and assembly of the data. Also, it may behoove those who manage such networks to implement data security and encryption measures to prevent the uninvited access of stored data by the end-user.

The invention may be conveniently implemented using a conventional general purpose digital computer and programmed according to the teachings of the present invention, as will be apparent to those skilled in the art. Appropriate software coding can readily be done by skilled programmers based on the teachings of the present disclosure, as will be apparent to those skilled in the software art.

The invention may also be implemented by the preparation of application specific integrated circuits or by interconnecting an appropriate network of conventional component circuits, as will be readily apparent to those skilled in the art.

The present invention also includes a computer program product which is a storage medium including instructions which can be used to program a computer to perform a process of the invention, such as the storage of the distributed programs, or the content flow control of the distribution network. The storage medium can include, without limitation, any type of disk including floppy disks, optical disks, CD-ROMs and magneto-optical disks, ROMs, RAMs, EPROMs, EEPROMs, magnetic or optical cards, or any type of media suitable for storing electronic instructions. As an apparatus embodying the present invention is operated, the device processes signals which are physical phenomena and control the device in a appropriate manner so that the desired results of the invention are achieved.

Obviously, numerous modifications and variations of the present invention are possible in light of the above teachings. It is therefore to be understood that within the scope of the appended claims, the invention may be practiced otherwise than as specifically described herein.

What is claimed is:

1. A data on-demand system for delivering data on-demand to a requesting end-user unit ("Requesting EUU") in a communications network, comprising:

a plurality of end-user units ("EUUs"), at least one of said EUU ("First Source EUU") adapted to receive a data file through said communications network, to store said data file and to transmit said data file through said communications network;

a server adapted to direct transmission of data files through said communications network among said plurality of EUUs,

wherein said server and said First Source EUU are adapted:

to permit said server to transmit said data file through said communications network to said First Source EUU;

to store said data file in said First Source EUU; and

to permit said server, in response to an instruction received from said Requesting EUU, to cause said First Source EUU to transmit said data file through said communications network to said Requesting EUU.

2. A data on-demand system as in Claim 1, wherein said server is additionally adapted to track information regarding the location of said data file at said First Source EUU.

3. A data on-demand system as in Claim 1, wherein said data file comprises at least one of the following:

an integral multimedia program;

at least one fragment of an integral multimedia program;

a combination of fragments of a plurality of integral multimedia programs.

4. A system according to Claim 1, further comprising a Second Source EUU, wherein:

said Second Source EUU is adapted to receive and store said data file as said data file is stored by said First Source EUU; and

5 said server and said Second Source EUU are adapted to cause said Second Source EUU to begin transmitting said data file to said Requesting EUU when transmission from said First Source EUU meets one or more of predetermined criteria.

5. A system according to Claim 1, wherein said server is adapted to cause said
10 First Source EUU to replace said data file in storage with at least one other data file based on at least one predetermined criterion.

6. A server for managing and delivering data files on-demand to a plurality of end-user units ("EUVs") connected through a data communications network, the server
15 comprising:

a central processing unit ("CPU");

a memory operatively connected to said CPU, said memory containing a program adapted to be executed by said CPU, wherein:

said CPU and said memory operatively adapted to distribute a plurality of data files
20 to said plurality of EUUs for storage based on at least one of predetermined criteria, while tracking the locations of said data files at said plurality of EUUs,

said CPU and said memory operatively adapted to receive a demand from a first EUU for at least one data file,

said CPU and said memory operatively adapted to locate said at least one data file
25 at a second EUU and to instruct said second EUU to transmit said at least one data file to said first EUU according to said demand.

7. A server according to Claim 6, wherein:

said CPU and said memory operatively adapted to receive a demand from a third EUU for a combination of a first data file and a second data file,

said CPU and said memory operatively adapted to locate said first data file stored at a fourth EUU and said second data file stored at a fifth EUU,

said CPU and said memory operatively adapted to instruct said fourth and fifth EUUs to transmit said first and second data files to said third EUU according to said demand.

8. A server according to Claim 7, wherein said combination of first and second data files demanded by said third EUU constitute 2 fragments from an integral data file.

9. A server according to Claim 7, wherein:

said CPU and said memory operatively adapted to locate and manage a plurality of data files and sequence transmission of said data files according to demands.

10. A server according to Claim 6, wherein:

said CPU and said memory operatively adapted to track usage information among said plurality of EUUs;

said CPU and said memory operatively adapted to re-distribute data files among said plurality of EUUs for storage in response to said usage information;

said CPU and said memory operatively adapted to add new data files for distribution among said plurality of EUUs for storage when said new data files are available.

11. An end-user unit for a data on-demand system, comprising:

a processor;

a modem adapted to connect with a data communications network to receive a data file, said modem also adapted to transmit at least a portion of said data file through
5 said data communication network to a second end-user unit specified by the server;

a storage unit adapted to store said data file in response to an instruction from a server connected to the data communication network.

12. A method of delivering at least one data file on-demand from a central
10 storage to a plurality of end-user units ("EUUs") through a communications network, comprising:

- a) storing at least one data file at said central storage;
- b) initializing a first EUU by transmitting at least a portion of said data file to said first EUU;
- 15 c) storing said portion of data file at said first EUU;
- d) tracking information regarding location of said portion of data file;
- e) determining location for said portion of data file, when a second EUU issues a demand for said data file;
- f) issuing instruction to said first EUU for said portion of data file stored at
20 said first EUU;
- g) upon receipt of said instruction, said first EUU transmitting said portion of data file to said second EUU.

13. The method according to Claim 12, the method further comprising:

- 25 a) initializing a third EUU by transmitting at least said portion of data file to a third EUU for storage;
- b) tracking information regarding location of said data file at said third EUU;
- c) monitoring transmission from said first EUU to said second EUU;
- d) if transmission from said first EUU to said second EUU falls below a
30 predetermined criterion, issuing instruction to said third EUU for

- transmission of said portion of data file to said second EUU and issuing instruction to said first EUU to terminate its transmission;
- e) upon receipt of said instruction, said third EUU transmitting said portion of data file to said second EUU.

5

14. A method of delivering at least one data file on-demand to at least one of a plurality of end-user units ("EUUs") through a communications network, comprising:

- a) providing a plurality of data files accessible through said communications network;
- b) distributing said plurality of data files to a plurality of EUUs for storage;
- c) tracking information regarding respective storage locations of said plurality of data files at said plurality of EUUs;
- d) when a requesting EUU issues a demand for at least one of said plurality of data files, locating at least one EUU storing said at least one of said plurality of data files from said plurality of EUUs based on at least one of predetermined criteria;
- e) causing the identified EUU to transmit said at least one data file to said requesting EUU.

20

15. The method according to Claim 14, wherein said predetermined criteria in the step d) of identifying comprises at least one of the following criteria:

- quality of transmission between each of said plurality of EUUs to said second EUU;
- proximity of each of said first plurality of EUUs relative to said second EUU;
- needs of each of said plurality of EUUs relative to a second plurality of EUUs.

30

16. The method according to Claim 14, wherein said plurality of data files comprises multimedia programs and data.

17. A method of delivering multimedia programs on-demand from a remote storage to a plurality of end-user units ("EUUs") through a communications network,
5 comprising the steps of:

- a) storing at least one multimedia program at said remote storage;
- b) initializing a first EEU by transmitting at least a first portion of said multimedia program to said first EEU;
- c) storing said first portion of said multimedia program at said first EEU;
- 10 d) initializing a second EEU by transmitting at least a second portion of said multimedia program to said second EEU;
- e) storing said second portion of said multimedia program at said second EEU;
- f) tracking information regarding storage locations of said first and second
15 portions of said multimedia program at said first EEU and second EEU;
- g) when a third EEU issues a demand for said multimedia program, directing said first and second portions of said multimedia program to be transmitted to said third EEU;
- h) upon receipt of said first and second portions of said multimedia program,
20 said third EEU assembling said multimedia program for downloading.

18. The method according to Claim 17, wherein said multimedia program comprises at least one of video programs, audio programs or movie programs.

25 19. The method according to Claim 17, wherein each portion of said multimedia program is encrypted with a predetermined encrypted key.

30 20. In a data communication network with a plurality of clients and servers, a method of delivering at least one data file from a plurality of data files on demand to a first client, comprising the steps of:

- a) distributing said plurality of data files to at least one of said plurality of clients and servers for storage based on a predetermined distribution scheme;
- b) tracking each of said data files as to its location;
- c) upon receiving a request for one of said plurality of data files from said first client;
- d) locating from at least one of said clients and said servers storing said requested data file;
- e) instructing at least one of said clients and said servers to transmit said requested data file to said first client.

10

21. The method according to claim 20, wherein said predetermined distribution scheme is based on at least one of: location of the clients and servers, and popularity of the data files.

15

22. A method of conducting a VOD service using a communications network, comprising the steps of:

- a) providing a plurality of movies which are accessible through at least one server in said communications network;
- b) distributing a plurality of set-top boxes ("STBs") to a plurality of clients, at least a portion of the STBs being capable of storing, presenting and transmitting said movies upon a predetermined request;
- c) storing the plurality of movies, in fragments or in whole, throughout said STBs and said at least one server, and keeping track of the locations of the fragments of said movies;
- d) upon request from one of the clients for a movie from said plurality of movies, locating said movie, in fragments or in whole;
- e) upon locating said movie, in fragments or in whole, among said STBs and said at least one server, causing said movie, in fragments or in whole, to be delivered to the requesting client.

25

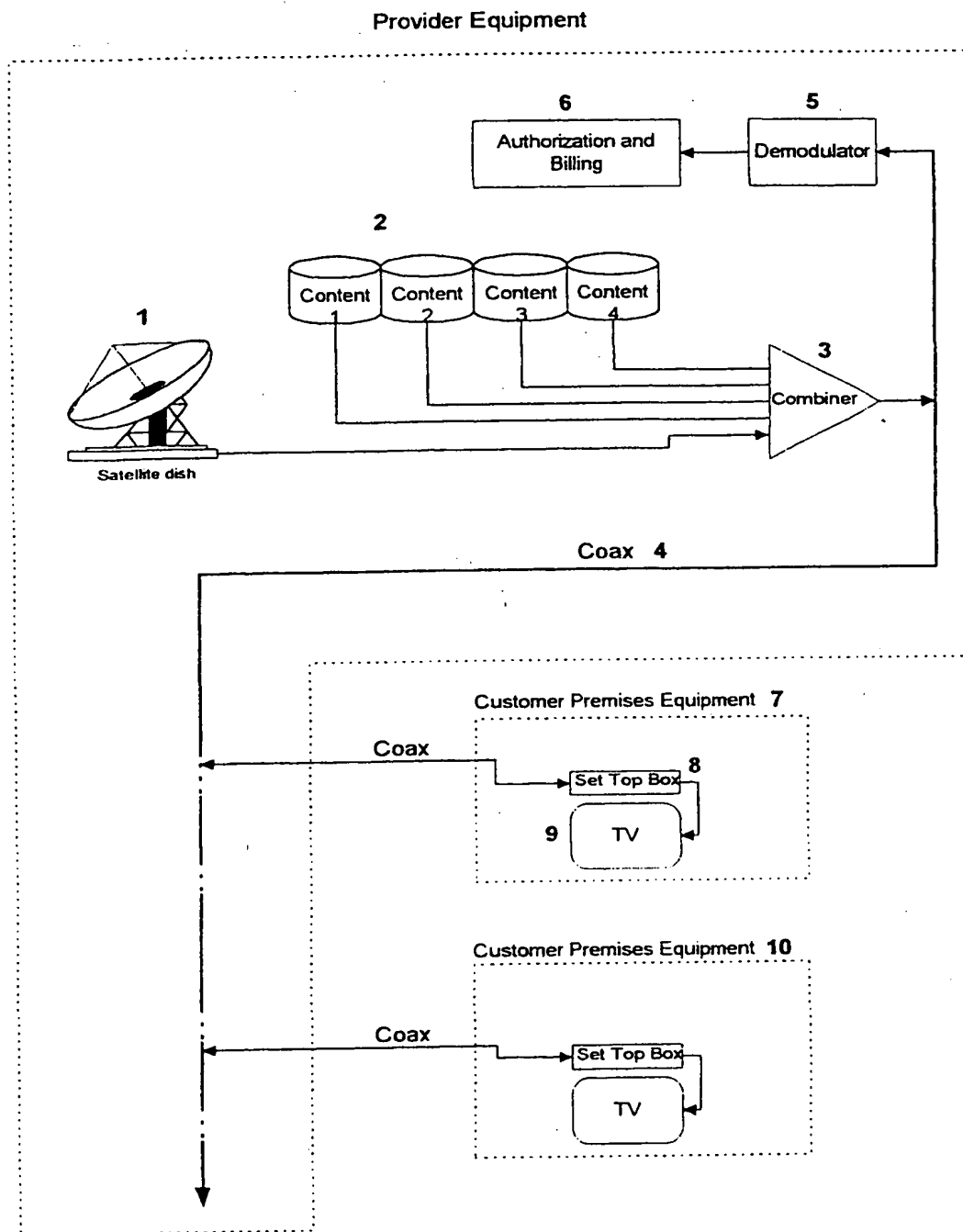


Figure 1A

2/11

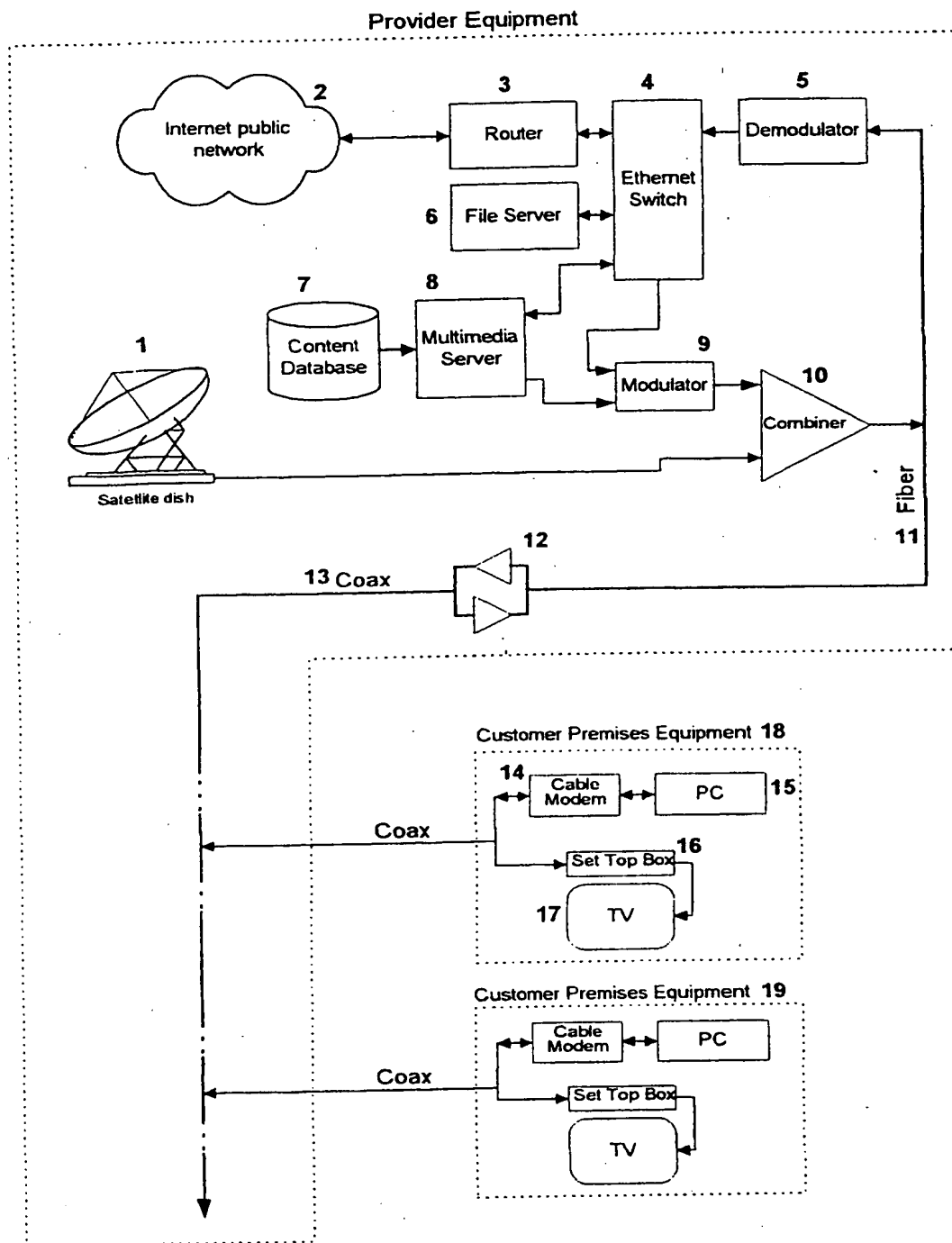


Figure 1B

3/11

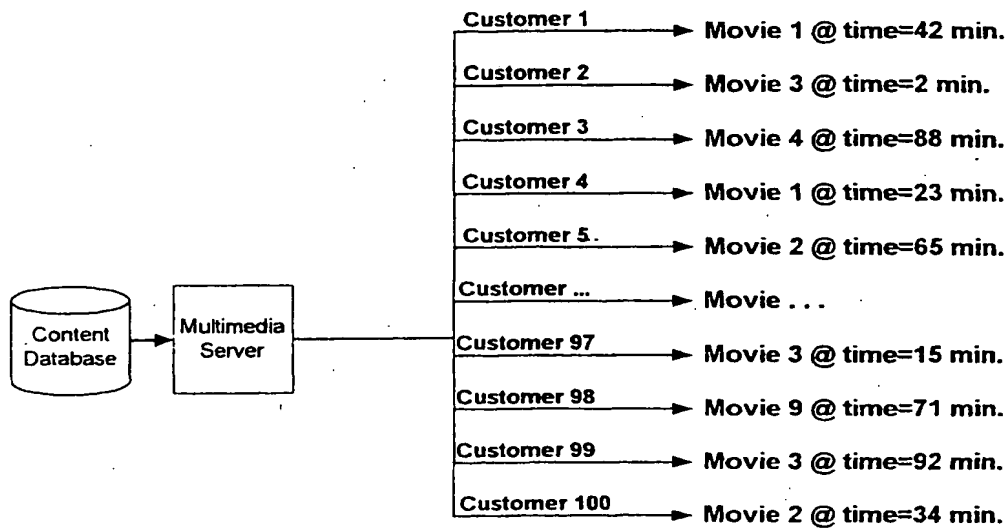


Figure 2A

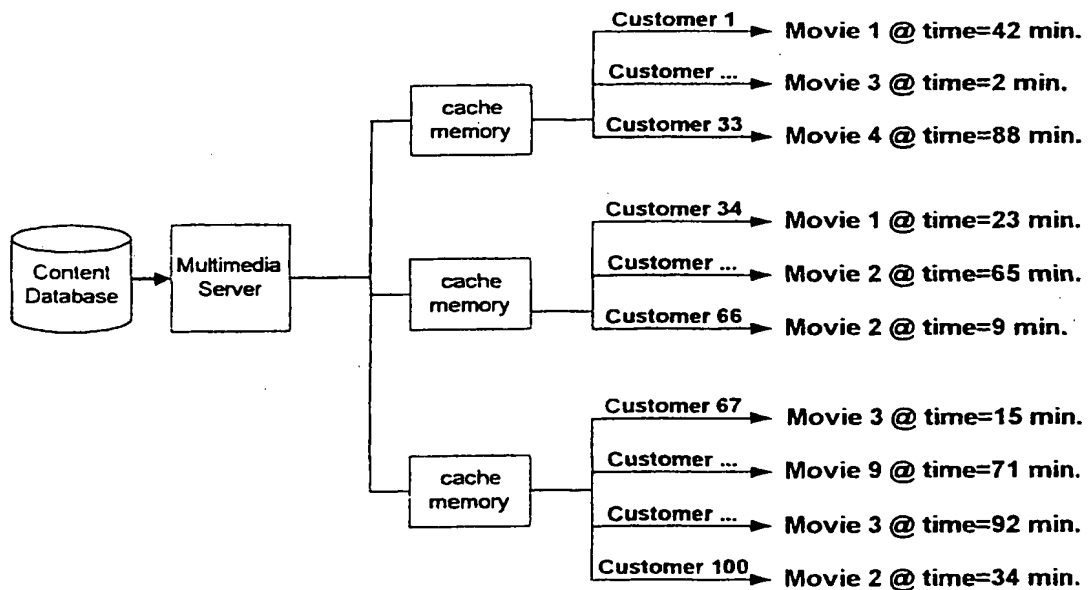


Figure 2B

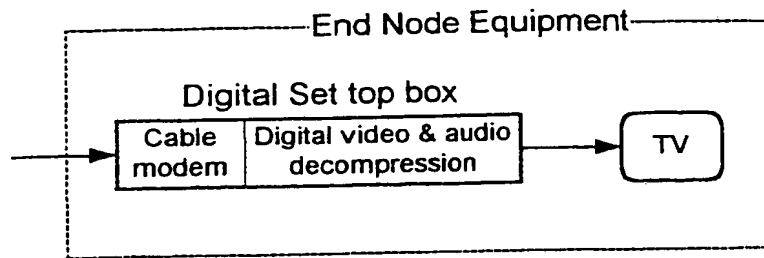


Figure 3A - Conventional End Node Equipment

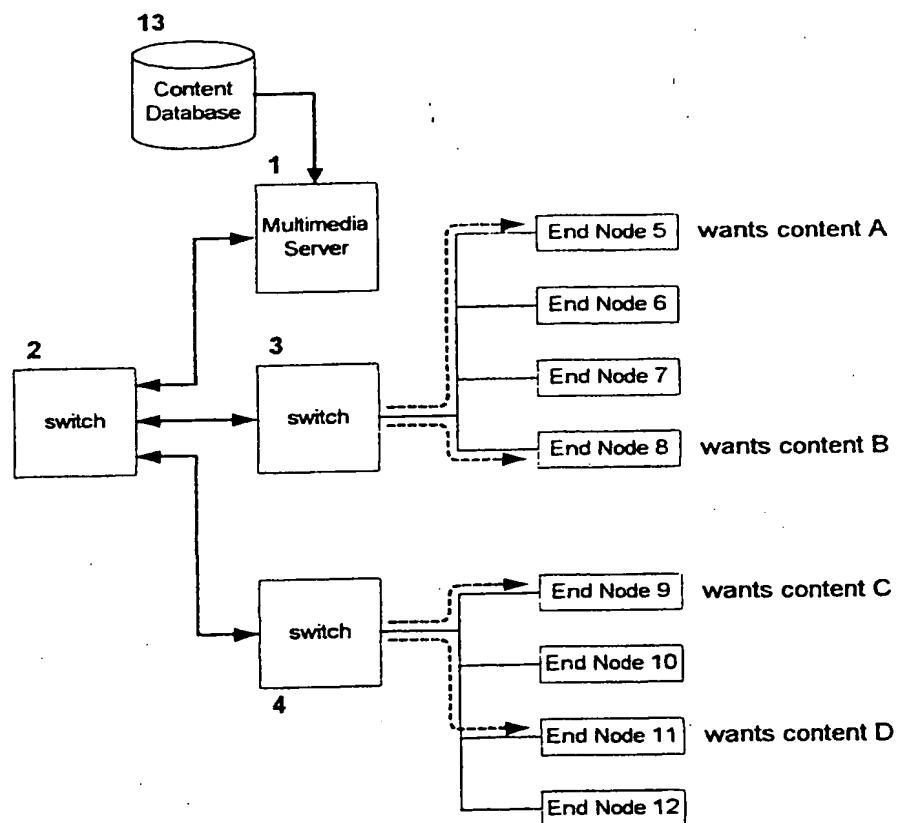


Figure 3B

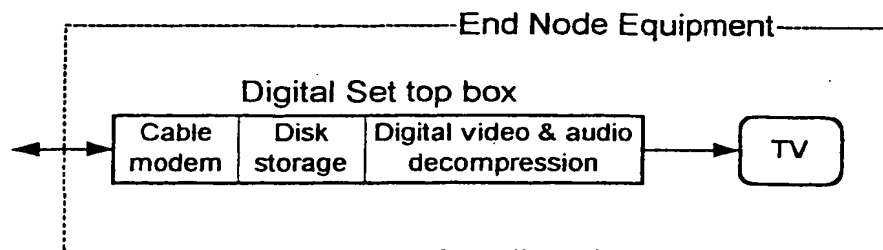


Figure 4A

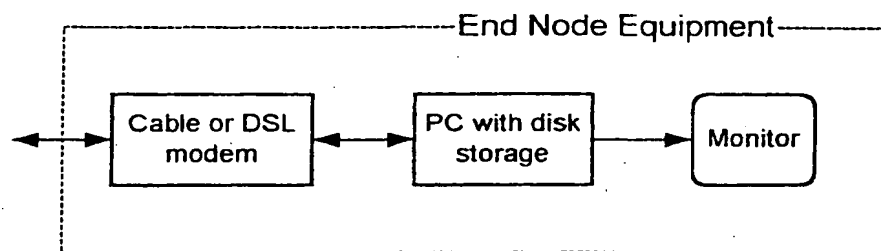


Figure 4B

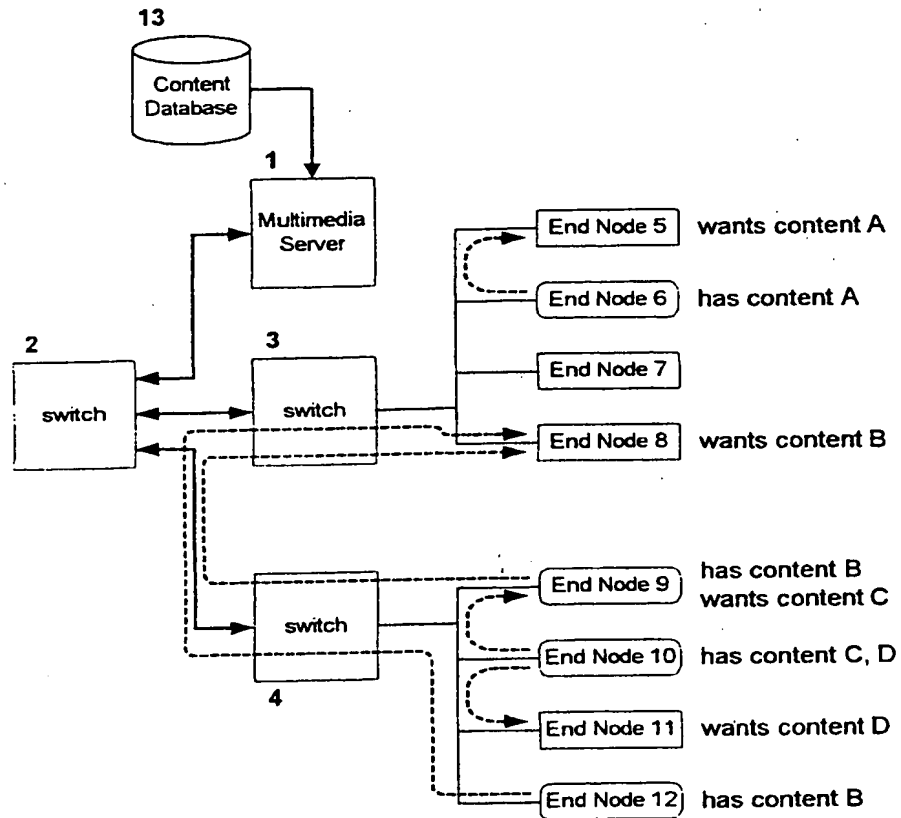


Figure 4C

7/11

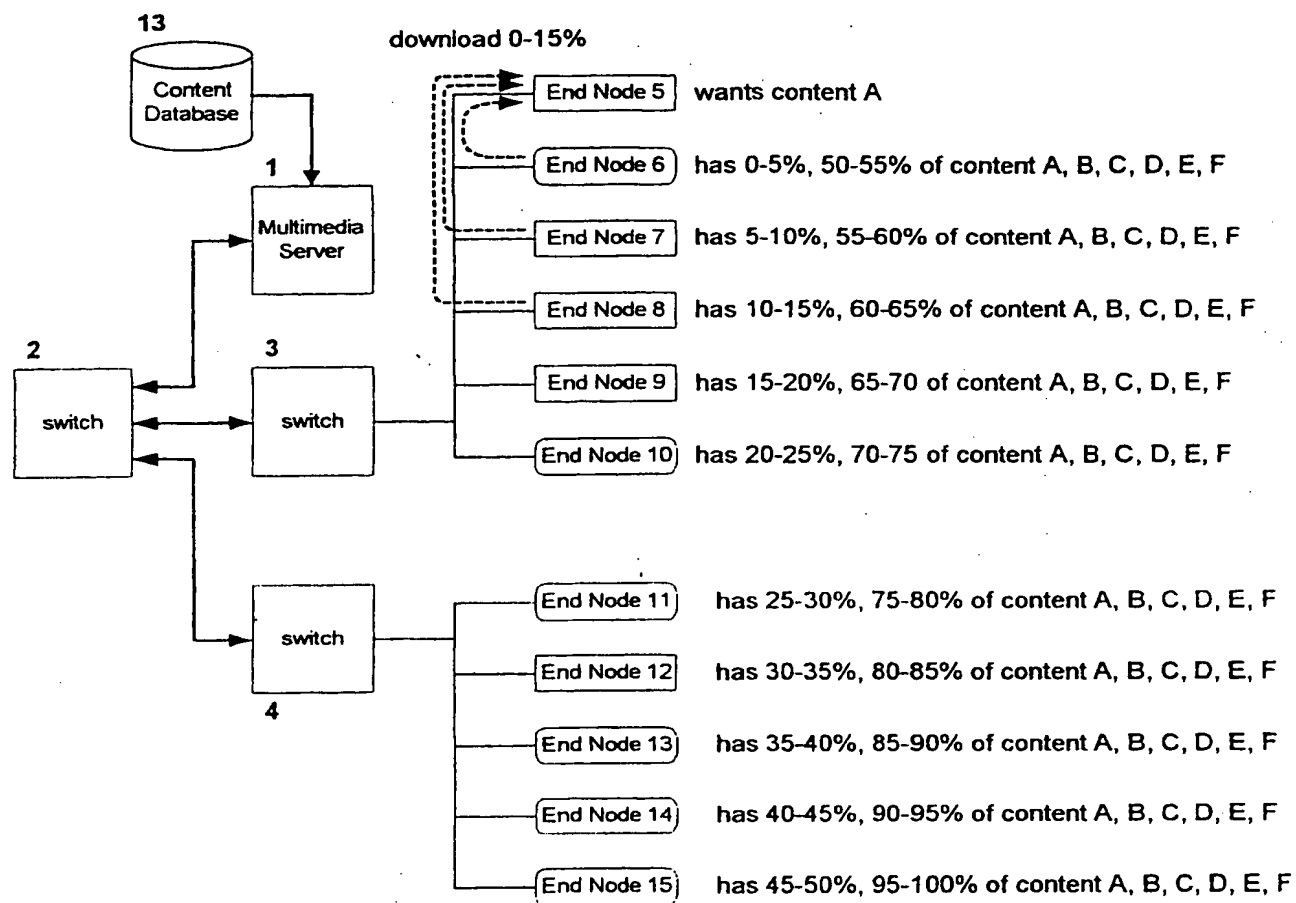
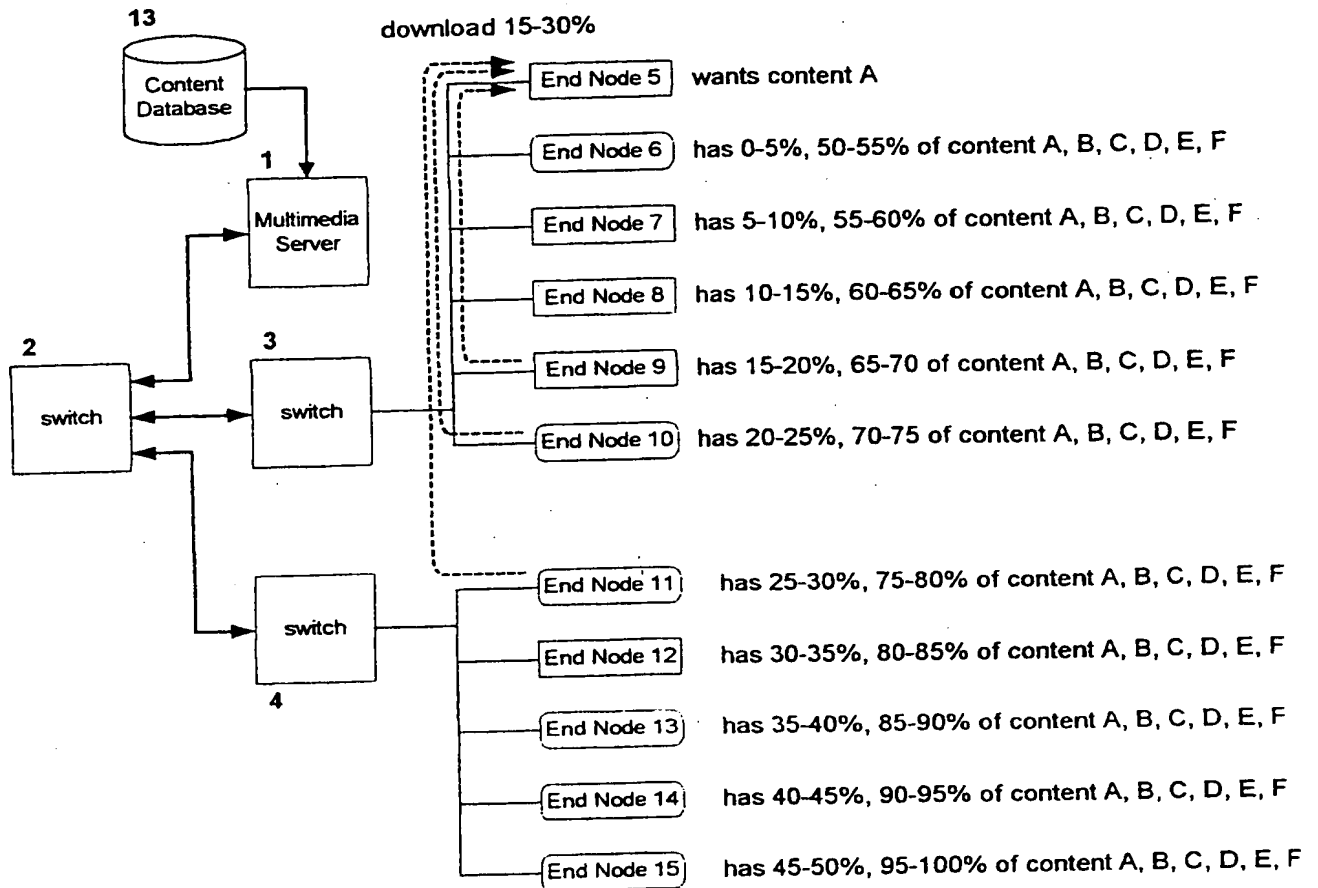
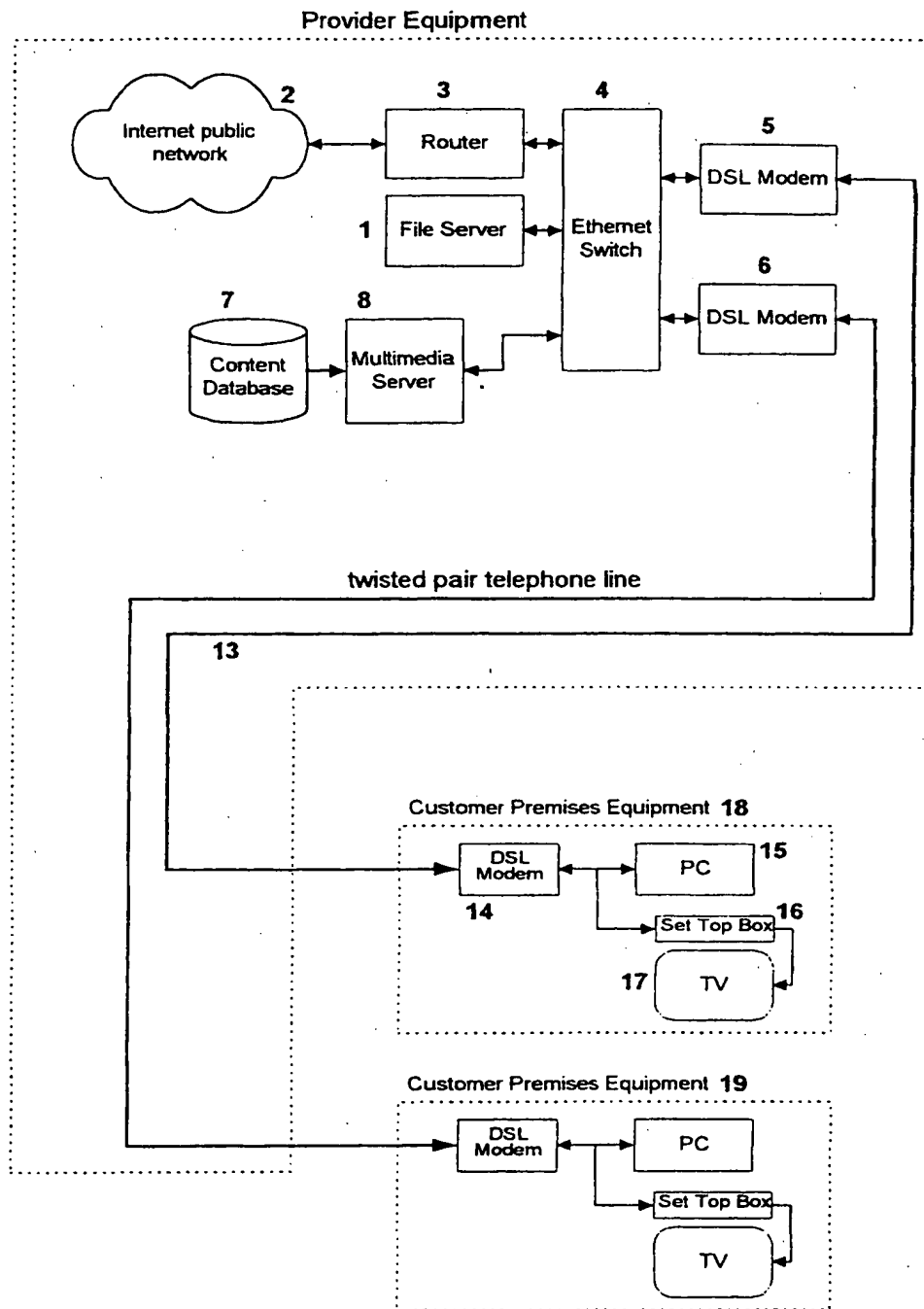


Figure 5A

**Figure 5B**

9/11

**Figure 6**

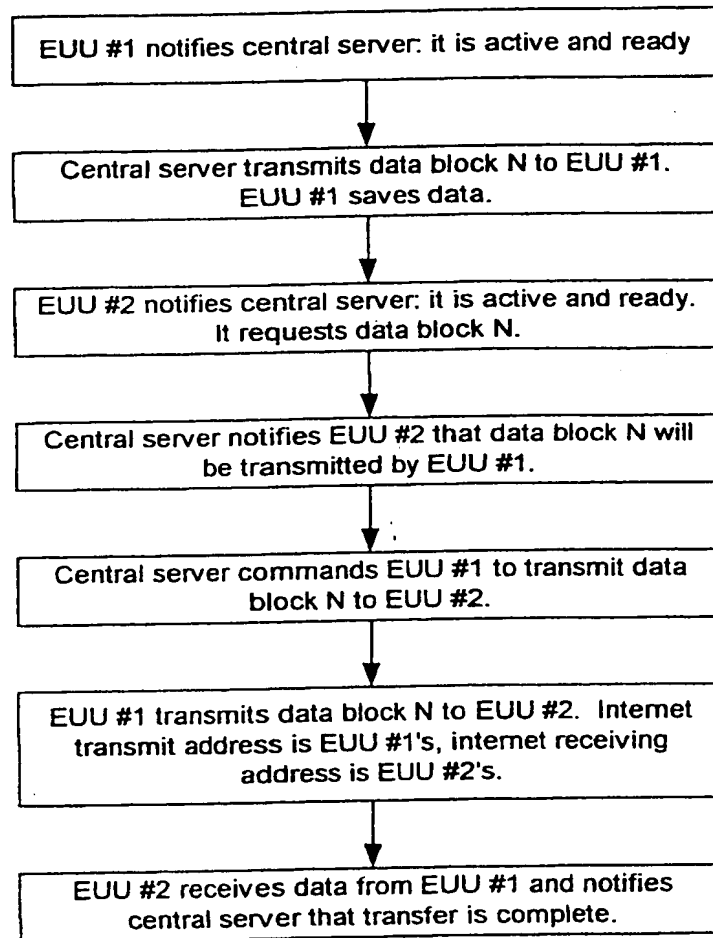


Figure 7

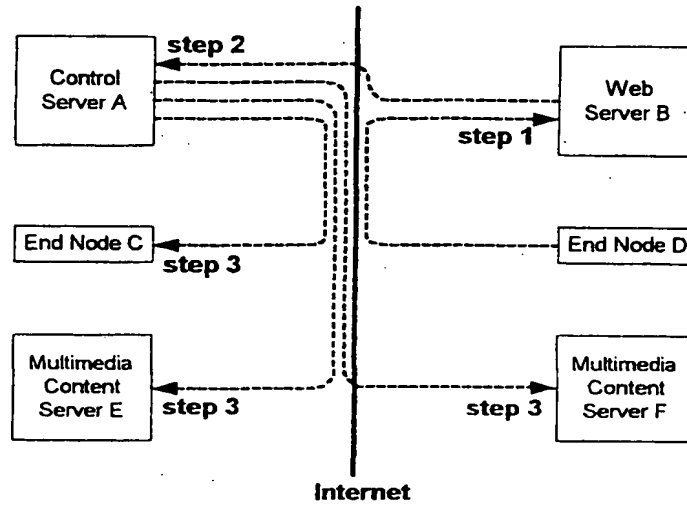


Figure 8A

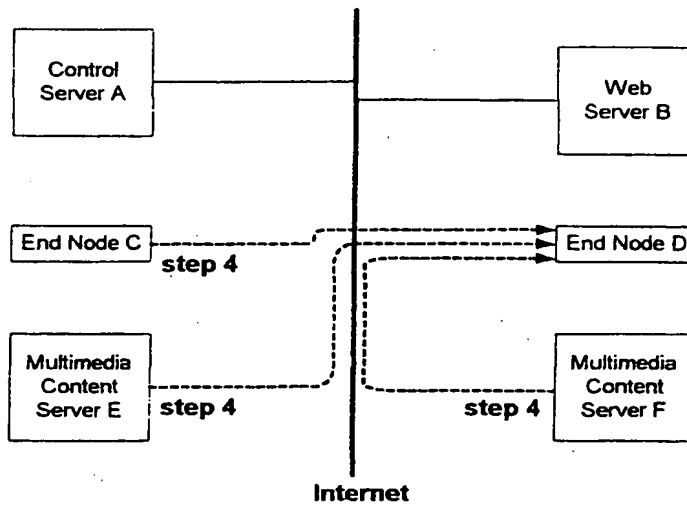


Figure 8B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/21127

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :H04N 7/173

US CL :709/219

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 709/219; 345/327; 348/7,12,13; 455/4.2, 5.1; H04N 7/173, 7/16

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,815,194 A (UEDA) 29 September 1998 whole document	1-22
A	US 5,873,022 A (HUIZER et al.) 16 February 1999 whole document	1-22
A	US 5,884,141 A (INOUE et al.) 16 March 1999 whole document	1-22



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

10 NOVEMBER 2000

Date of mailing of the international search report

08 JAN 2001

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

CHRISTOPHER GRANT

Telephone No. (703) 305-4755